Augmenting Camera Images for Operators of Unmanned Aerial Vehicles

Hans (J.A.) Veltman and Arjen (A.B.) Oving
TNO Human Factors
Kampweg 5
P.O. Box 23
3769 ZG Soesterberg
THE NETHERLANDS

SUMMARY

The manual control of the camera of an unmanned aerial vehicle (UAV) can be difficult due to several factors such as 1) time delays between steering input and changes of the monitor content, 2) low update rates of the camera images and 3) lack of situation awareness due to the remote position of the operator and the small field of view of the camera images. Therefore, it is important to assist the operator with adequate tools. We constructed and evaluated a perspective (3D) digital map with information about the predicted viewing direction of the camera. In an experiment, participants inspected roads and edges of wood with and without the 3D map. A 2D map with the same information as the 3D map was available in all conditions. With the 3D map, the participants: 1) were able to inspect larger areas, especially when the task became more difficult due to time delays and low update rates, 2) were better able to perform an additional task, and 3) reported lower workload compared to the condition without the 3D map. These subjective workload results were not supported by objective (physiological) workload measures. Analysis of eye movements showed that the 3D map was used very frequently, especially in conditions with time delays and low update rates.

1.0 INTRODUCTION

Unmanned aerial vehicles (UAVs) do not have pilots on board, but operators are very important for the control of the system during operation. Due to a high level of automation, steering of the platform has become relatively easy and most often UAVs follow pre-programmed waypoints. However, manual *camera* steering is often required. To control the camera, a steering signal from a control station is sent to the camera. The camera then turns to the required position and the camera images are sent back to the control station. In many cases the images also must be coded and decoded. This loop may cause substantial *time delays* between the control input and changes of the camera pictures in the control room. Time delays between steering input and display of the camera images may seriously hamper task performance. For instance, a delay of 0.3 s already results in a substantial performance decrement. This is because the operator has to wait for the result of the steering input before the next input can be generated (Cooke, 1965; Wagenaar, 1971; Veldhuyzen & Stassen, 1973). Fast feedback about the results of the steering input is therefore essential for adequate steering performance. In addition, *update frequency* of the images can be low due to bandwidth limitations. Very low update rates will lead to time delays (with a 3 Hz update rate, the time delay will be at least 0.33 s) and a snapshot like presentation of images, without any perceptual information on motion. Another problem is a *reduced situation awareness* of the operator. This is because 1) the cameras often have a small field of view and 2) the viewing direction of the camera is not the same as the flight direction of the aircraft. Therefore, it can be difficult for the operator to interpret which part of the environment is actually viewed and where the

Paper presented at the RTO HFM Symposium on "The Role of Humans in Intelligent and Automated Systems", held in Warsaw, Poland, 7-9 October 2002, and published in RTO-MP-088.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
00 OCT 2003	**N/A**	**-**

4. TITLE AND SUBTITLE	
Augmenting Camera Images for Operators of Unmanned Aerial Vehicles	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **TNO Human Factors Kampweg 5 P.O. Box 23 3769 ZG Soesterberg THE NETHERLANDS**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S)
	11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM001577., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**10**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

camera must be directed to. These factors make the control of a camera difficult and therefore, a more optimal design of the control interface can improve the performance and effectiveness of the UAV system.

TNO Human Factors conducts experiments on UAV control for more than a decade. One of the results is that problems with time delays and low update rates can be compensated for by using a computer-generated grid on or around the camera image (see Figure 1). This grid is a 'virtual' world that is drawn from the viewpoint of the camera and moves with the same speed as the camera, but it does not have a time delay between the steering input and the movement. The grid can be generated easily in the control station, by using the control input of the operator directly. Therefore, the movements of the grid are not affected by the time delays involved in sending and retrieving the UAV data. In addition, the update rate of the grid can be set sufficiently high for adequate motion perception. When there is a time delay in the camera images, the grid will move sooner than the camera images, which makes steering much easier due to the direct feedback on the magnitude of the control input. Information presented by such a grid results in improved task performance (Korteling, Van Erp & Kappé, 1997). Besides control feedback, preview is important as well for good steering performance. Especially when the camera is zoomed in to look at a small part of the environment, preview is very limited. Following a moving object on a road, for example, can be very difficult because the operator does not have information about the direction of the road in front of the object, and therefore, the operator can not anticipate adequately.

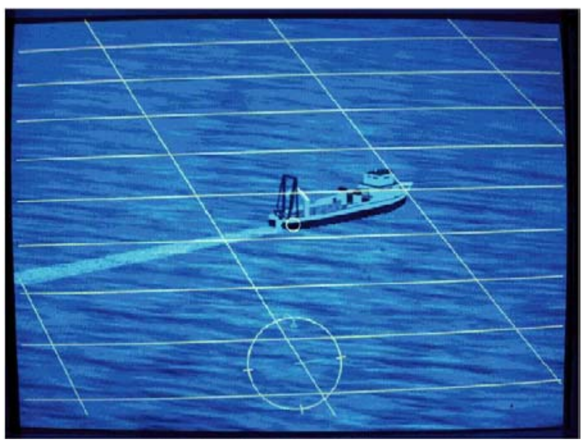

Figure 1: A Simulated Camera View of a Ship at Sea with the Computer-Generated Grid as an Overlay. The orientation of the grid adheres to the perspective geometry of the camera viewpoint to provide accurate motion feedback on control input.

Camera images are often supported by digital 2D maps showing the position of the UAV and the part of the environment at which the camera is looking (footprint). A digital map with this information increases situation awareness of the operator and provides preview, which makes steering of the camera easier. However, problems with time delays and low update rates are not solved with a footprint. These problems can be reduced by using a second footprint that shows the predicted camera position (Van Erp & Kappé, 1998). The principle is similar to the grid that has been described above. From a human factors perspective, a 2D map is not always optimal for camera control. A 2D map can be used when *global* information is required, especially when the map is presented north-up. A 3D map is more optimal for tasks for which *local* information is required (Wickens & Prevett; 1995). Global information refers to the position of objects relative to other objects in an earth reference frame (e.g., the position of a bridge relative to a city). Local information refers to the position of objects relative to the viewing position of the camera (e.g., a road ahead of a vehicle that has to be tracked).

TNO Human Factors developed a simulation environment in which human factors principles for UAV camera control can be demonstrated and in which experimental studies can be conducted. In the present experiment we used this simulator environment to investigate the benefits of a 3D map with regard to operator performance and mental workload. We constructed a 2D map (oriented north-up) in which feedback about the steering input was provided by means of an additional footprint that showed the predicted viewpoint of the camera. Operators had to find targets on roads and along wood edges. They could use the map to see to which part of the environment the camera was oriented. In one half of the conditions a 3D map was available together with the 2D map. The 3D map showed the same information, but was drawn from the viewpoint of the camera. In some conditions, the quality of the camera images was manipulated by introducing a time delay of 1 second or by lowering the update rate of the camera images to 3Hz. This had no effect on the 2D and 3D map. Furthermore, in one half of the conditions a secondary task had to be performed. This was done to see whether operators had more spare capacity when the 3D map was available. Several performance measures were distracted and workload was measured with subjective and physiological measures.

2.0 METHOD

2.1 Simulator

The simulator environment consisted of two 21″ displays (see Figure 2) that were each connected to a graphics computer (SimFusion, Evans & Sutherland). The resolution of each display was 1280 by 1024 pixels. The simulated camera could be turned 360^o horizontally and 90^o vertically by using a joystick. The zoom factor could be adjusted up to 20 time the original size by using a button on the joystick. The camera images could be shown in three modes: 1) *normal*: low time delay (0.1 s) and between 10 and 30 Hz update rate (depending on the amount information in the database), 2) *low update rate* (3 Hz, in combination with the normal time delay) and 3) *large time delay* (1 s, in combination with the normal update rate).

Figure 2: The Left Panel shows the 2D Map with the Position of the UAV in the Center. The right panel shows the 3d map, which is drawn from the viewpoint of the camera. In both displays, the yellow footprint shows the part of the environment that corresponds to the camera image and the orange footprint shows the predicted position. The camera images are presented at the bottom of the right display.

2D Map

The left panel of the simulator displayed a detailed 2D map of the environment (see Figure 2, left side). This map was presented north-up. Apart from terrain information (roads, woods, buildings etc.) it contained the following relevant information: waypoints and the route of the UAV were shown (yellow line), flight direction of the UAV (orange arrow) and the actual and predicted footprints (see 'Footprint' below for an explanation). The 2D map was a virtual 3D world that was viewed from above at an altitude distance of 6500 m. The fixed field of view was 60^o (H) by 48^o (V) showing an area of 7.2 by 5.6 km.

3D Map

The right panel of the simulator displayed the 3D map (see Figure 2, right side). In this map, the virtual 3D world was drawn from the viewpoint of the camera. The UAV flew at a fixed altitude of 1200 m. The 3D map also had a fixed field of view of 60^o (H) by 48^o (V). Terrain information was better visible in this 3D map because the altitude of the UAV was less than the virtual camera position used for the 2D map. The viewpoint for generating the 3D map depended on the camera control input of the operator, and the angular motion of this viewpoint was identical to the angular motion of the camera. Furthermore, because the vertical viewing angle was most often less than 90^o (i.e., when looking straight down), a perspective view of the database environment was obtained.

Camera Images

The simulated camera images were presented in the lower center of the right panel and was sized 1/9 of the display (resolution 427 by 341 pixels).

Footprints

Both the 2D and 3D map had two footprints. The yellow footprint showed the section of the map that corresponded to the camera images. The orange footprint provided direct feedback about the steering input. It was directly linked to the stick input and thus, provided predicted information about the camera position. In the conditions with low update rates and time delays, the yellow footprint followed the orange footprint. The size of the footprint could be adjusted by the zoom function, providing feedback about the zoom setting. When the zoom factor was 1, the size of the footprint on the right display was the same as the size of the camera panel. In the conditions in which the 3D maps were not drawn on the right display, the footprints were still visible to provide feedback about the zoom settings.

It should be noted that in real situations, the following factors are responsible for time delays between steering input and changes of the camera images: 1) time needed for sending information from the control station to the UAV, 2) response time of the camera, 3) time needed for coding, sending, and decoding information from the UAV to the control station. For the present experiment, we only simulated the last factor because this is easier to implement in a simulator. Discrepancies between a real situation and the simulated situation occur when there is a time delay from the control station to the UAV. The discrepancy depends on the altitude and speed of the UAV. Higher speed and lower altitude will result in a larger discrepancy in viewing direction of the footprint. When the delay is 0.5 s and the UAV flies at an altitude of 1200 m with an airspeed of 120 km/h (as used in the present experiment), the footprint is drawn from a viewing position that differs 1.5 degrees from the correct viewing position. This does not result in noticeable differences for the operator.

2.2 Tasks

Primary Task

The primary task was to search and detect military vehicles. The simulated UAV flew a planned route above a small village that was surrounded by several parcels of wood. Several military vehicles were positioned along the wood edges (Leopard II, ZSU-24, YPR and T72). The participants had to detect as many Leopard II and ZSU-24 tanks by pressing a button on the joystick. The other tanks (YPR and T72) had similarities to the other tanks with respect to color and shape. Therefore, it was important to look carefully before giving a response. The camera had to be steered along the wood edges and along the roads. This is a reconnaissance task, for which local information is supposed to be most optimal, because the camera has to be navigated along the wood edges.

Secondary Task

The secondary task was a reaction and memory task. Two white squares were presented above each other. Every 10 s one of these squares became red or blue. A blue square contained a number "1" or "2" indicating the number of time this square had been red before. The participants had to press a 'yes' or 'no' button on the joystick after a blue square appeared, indicating whether the number was correct or not. For an adequate performance, participants had to inspect this display regularly. Furthermore, this task was mentally demanding because changing information had to be kept continuously in working memory. Furthermore, the task was presented on a small display that was placed in the middle below the two simulator displays.

2.3 Experimental Factors

The following factors were manipulated in the experiment:

- *3D map* (2 levels): with and without 3D map. In the conditions without the 3D map, the right panel only contained the camera images and the two footprints that provided feedback about the steering input and zoom factor. The 2D map was present in all conditions.
- *Quality of camera images* (3 levels): normal, 3Hz-update rate and 1 s time delay.
- *Secondary task* (2 levels): with and without secondary task.

A complete combination of these three factors resulted in 12 conditions (3 x 2 x 2), that were all performed by the participants. Each condition lasted seven minutes. The data were analyzed according to an ANOVA repeated measurement design.

2.4 Participants

Twelve high school students participated in the experiment (average age 17.6 years). Furthermore, two experienced UAV sensor operators of the Royal Netherlands Army participated (29 and 31 years). The order of the conditions was completely balanced with twelve students. With all fourteen participants, only the factor '3D map' was completely balanced. Therefore, the data were analyzed for twelve students and for the complete group of fourteen participants. The results did not differ and therefore, the results from the complete group will be presented. The participants came in pairs to the laboratory for one day. During the morning all conditions were trained and the experiment took place in the afternoon in which the participants performed blocks of four conditions after each other.

2.5 Measures

The following performance measures were derived: 1) percentage of woods edges and roads that were inspected, 2) percentage of vehicles that were detected correctly, 3) performance on the secondary task (percentage correct and reaction times), 4) stick control: standard deviation of the XY-position of the stick, and applied zoom factor. The following workload measures were used: 1) subjective effort rating after each condition (Rating scale Subjective Mental Effort, RSME; Zijlstra, 1993) and 2) physiological measures (heart rate, heart rate variability, respiration and blood pressure). Furthermore, the electroencephalogram (EOG) was measured to distract eye blinks and fast eye movements.

3.0 RESULTS

Searching

The analysis of the percentage of inspected wood edges showed main effects of the factors 3D map and Quality of camera images. More importantly, however, was the significant interaction between these two factors [$F(2,26)=4.8$, $p<.05$]. Figure 3 shows the percentage of wood edges that are inspected as a function of 3D map and camera quality. About 35% of the edges of the woods were inspected when the quality of the camera images was normal. The 3D map resulted in a very small increase in performance. In the condition with a lower quality of the camera images (3 Hz update rate and 1 s time delay) the percentage dropped to about 20% when no 3D map was available. However, with the 3D map about 30% was inspected, which is about 50% better. These results show that the performance decrement due to a low quality of the camera images can be partially compensated for by providing a 3D map.

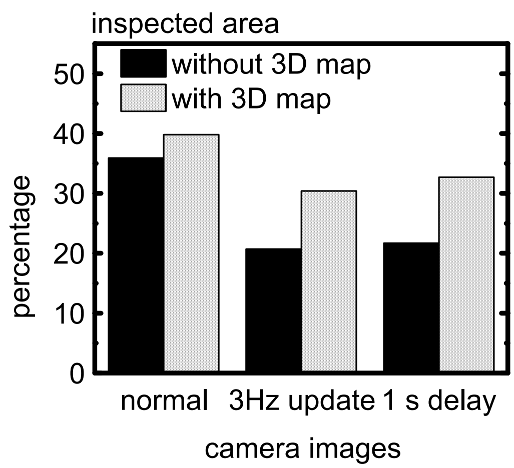

Figure 3: Percentage of Inspected Wood Edges as a Function of 3D Map and the Quality of Camera Images.

Despite the larger area that was inspected when the participants had a 3D map, they did not find more targets. Further analysis of the data revealed that this was due to the distribution of the targets along the edges of the woods. In each condition the UAV flew a different route over the village. The target locations were different in each condition. Most targets were positioned close to the routes of the UAV. The participants did not know this in advance and expected the same target probability at each location. The extra edges of the woods that were inspected in the conditions with the 3D map were mainly far away from the route of the UAV. Thus, they inspected locations at which far less targets were present.

Stick Control

A slightly higher average zoom factor was used in the conditions with the 3D map (5.7 versus 5.1; $F(1,13)=3.9$, $p=.06$). Furthermore, the standard deviation of the stick input was smaller with the 3D map indicating that the participants moved the camera smoother (16% versus 17%; $F(1,13)=6.0$, $p<.05$). Detection of targets is easier with a higher zoom factor. However, the steering of the camera becomes more difficult because the operator has less preview and objects move faster on the display. Nevertheless, they were able to move the camera smoother. These results indicate that the participants were indeed better supported in steering the camera when they had a 3D map.

Secondary Task

The average percentage of correct responses was 67% while the chance probability for this task is 50%. Thus, the secondary task performance was very low. The secondary task did not affect the performance on the main task. It seemed that the participants tried to do their best for the main task and did not have much spare capacity for performing the secondary task. The performance on the secondary task was better when the 3D map was available (70% versus 64% correct; $F(1,13)=5.9$, $p<.05$). This is striking because the participants had three displays with information in the 3D-map conditions and only 2 displays in the conditions without the 3D map. Thus, the result of the 3D map was that the participants had more spare capacity for conducting additional tasks.

Mental Workload

Figure 4 presents the subjective effort ratings for the conditions with and without the 3D map and for the three camera conditions (note: higher ratings indicate more subjective effort). The effort ratings were lower when the 3D map was present [$F(1,13)=6.5$, $p<.05$]. The quality of the camera images also affected the subjective effort ratings [$F(2,26)=12.6$, $p<.001$]. The highest ratings were obtained for the conditions with 1 s time delay. In the conditions with the secondary task the effort rating was significantly higher (55 versus 44; $F(1,13)=16.3$, $p<.01$). No significant interactions were found for the subjective effort data. The physiological effort measures did not change as a function of quality of the camera images and 3D map. There was only a small difference in operator state as a function of the secondary task. Heart rate was increased slightly in the conditions with the secondary tasks (72.8 versus 72.0; $F(1,13)=9.0$, $p<.05$).

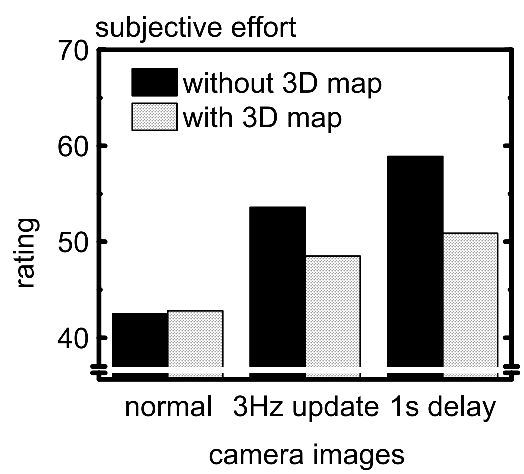

Figure 4: Subjective Effort Rating as a Function of Quality of the Camera Images and 3D Map.

Eye Movements

The frequency of eye blinks was the same in the conditions with and without the 3D map (about 10 blinks per minute), indicating that the visual load was the same in both conditions. Figure 5 shows that the number of vertical eye movements varied considerably as a function of the 3D map ($F(1,13)=27.6$, $p<.001$). Furthermore, eye movement frequency changed as function of the quality of the camera images ($F(2,26)=26.9$, $p<.001$), but only for the conditions with 3D map (interaction effect; $F(2,26)=6.7$, $p<.05$). These results show that the participants often inspected the 3D map, especially when the quality of the camera images was low.

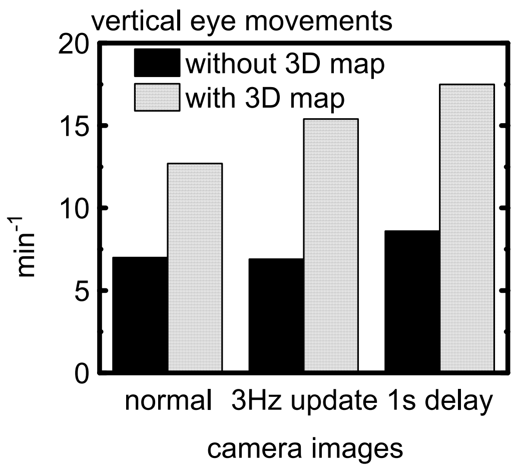

Figure 5: Frequency of Vertical Eye Movements as a Function of Quality of the Camera Images and 3D Map.

4.0 DISCUSSION

The 3D map resulted in an improvement of task performance. A larger area was inspected, performance on secondary task was improved indicating that the participants had more spare capacity and the participants reported lower effort. The positive effect of the 3D map was largest when the quality of the camera images was low. It is important to note that the adequate feedback about time delays and low update rates was always available in both the 2D and 3D map display. Without this information, performance would be much worse in the conditions with low update rates and time delay.

The subjective effort measure showed substantial effects as function of all experimental factors, whereas only heart rate showed a small effect as function of the secondary task. This can be due to the lower sensitivity of physiological measures for mental effort. However, we do find such discrepancies between subjective and physiological measures often (Veltman, Gaillard & Van Breda, 1997). This is related to the type of task that is evaluated. Participants most often give higher effort ratings when a task becomes more difficult as a result of a reduced quality of information. Physiological measures most often do not show differences in these situations, because investing more effort most often will not result in better task performance. When an additional task has to be performed, then attention has to be divided between more tasks. In these situations, additional effort has to be invested in order to keep an adequate level of performance of the main task. This is reflected in both subjective and physiological measures. The difference between the subjective and physiological measures can be explained along this line of thought. Degrading the quality of the display makes the task more difficult, resulting in higher effort ratings, but does not affect physiological measures because investing more effort will

not improve performance. For the secondary task, more effort was required in order to maintain an adequate level of performance on the main task.

There were not only positive effects of the 3D map. Participants inspected more edges of the woods far away from the UAV and used a higher zoom factor when the 3D map was present. It can be argued that this is a positive effect of the 3D map, making is possible to search at greater distances. However, the participants knew that the time available was not enough to search the complete area. Therefore, they had to decide from time to time whether they should search a complete parcel of wood or skip to a new parcel that was closer to the present location of the UAV. With the 3D map, some participants had the tendency to search the complete parcel of wood before they moved the camera to a new location. This might be due to the information about the distance between UAV and the part of the environment where the participant was looking at. The 2D provided much better feedback about this distance than the 3D map. When the participant used mainly the 3D map, it was more difficult to decide when they had to move to a new location. To overcome this problem, better information about the distance must be provided in the 3D map.

An additional display was used in the present experiment for the 3D map. The 2D map was always present. Each display has advantages and disadvantages, depending on the type of task. The 2D display provided global information and the 3D display provided local information. Using two different displays is not an optimal situation for a control station, where lack of space is often a problem. A better solution will be the integration of a 2D and 3D display. This can be achieved by using an adaptable display in which the operator can choose the presentation that is most appropriate. When the operator can adjust the viewing point from which the map is drawn, many different map presentation can be achieved. De Vries, Veltman and Van Breda (1999) used such an adaptable cockpit display in a high-speed flight task. Pilots could select different view points such as 2D north up, 2D heading up, 3D map drawn from the viewing position of the cockpit and a 3D map that was drawn from behind the own aircraft to present a larger area.

A 3D display may also be used to use other ways of steering the camera. A 3D display that is drawn from the viewpoint of the camera can be used to steer the camera by using a mouse or touch screen. This can be achieved by pointing to a location on the map, after which the camera moves to that position.

Conclusion

Using a 3D map for UAV camera control improves task performance for reconnaissance tasks, especially when the quality of the camera images is low.

5.0 REFERENCES

Cooke, J.E. (1965). *Human Decision in the Control of a Slow Response System.* Thesis, University of Oxford.

De Vries, S.C., Veltman, J.A. & Van Breda, L. (1999). *Use of Adaptable Displays for Fighter Aircraft Flight Support.* TNO Rapport (TNO-TM 1999 A011). Soesterberg, The Netherlands: TNO Human Factors Research Institute.

Korteling, J.E., van Erp, J.B.F. & Kappé, B. (1997). Visual Support for the Control of Unmanned Platforms. In D. Harris (ed.) *Engineering Psychology and Cognitive Ergonomics* (pp. 55-63). Brookfield: Ashgate.

van Erp, J.B.F. & Kappé, B. (1998). *Effects of Low Update Rate, Time Delay, and Footprint Prediction on Camera Control from Unmanned Systems.* Report TM-98-A003. Soesterberg, The Netherlands: TNO Human Factors Research Institute.

Veldhuyzen, W. & Stassen, H.G. (1973). Modelling the Behaviour of the Helmsman Steering a Ship. *AMC (9),* Cambridge, MA.

Veltman, J.A., Gaillard, A.W.K. & Van Breda, L. (1997). Workload Indices: Physiological Measures versus Subjective Ratings. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics. Volume One: Transportation Systems.* (pp. 269-275). Aldershot, Ashgate.

Wagenaar, W.A. (1971). *The Helmsman as an Adaptive Controller.* (Report IZF 1971 C-14). Soesterberg, The Netherlands: TNO Institute for Perception.

Wickens, C.D. & Prevett, T.T. (1995). Exploring the Dimensions of Egocentricity in Aircraft Navigation Displays. *Journal of Experimental Psychology.* Vol. 1, No. 2, pp.110-135.

Zijlstra, F.R.H. (1993). *Efficiency in Work Behaviour: A Design Approach for Modern Tools.* Thesis. Technical University of Delft.